Are cognitive differences between immigrant and majority groups diminishing?

te Nijenhuis, J.; de Jong, M.J.; Evers, A.; van der Flier, H.

published in
European Journal of Personality
2004

DOI (link to publisher)
10.1002/per.511

document version
Publisher's PDF, also known as Version of record

European Journal of Personality
Eur. J. Pers. **18**: 405–434 (2004)
Published online 14 May 2004 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/per.511

Are Cognitive Differences Between Immigrant and Majority Groups Diminishing?

JAN TE NIJENHUIS[1]*, MART-JAN DE JONG[2], ARNE EVERS[3] and HENK VAN DER FLIER[4]

[1]*Social and Organizational Psychology, Leiden University, The Netherlands*
[2]*Sociology, Erasmus University, Rotterdam, The Netherlands*
[3]*Work and Organizational Psychology, University of Amsterdam, The Netherlands*
[4]*Work and Organizational Psychology, Vrije Universiteit, Amsterdam, The Netherlands*

Abstract

A review is given of scores on various cognitive measures, comparing groups of ethnic Dutch and non-Western immigrants using a large number of datasets.

The research shows that there are large group differences in school results, work proficiency, and g *for Turks, Moroccans, Surinamese, Netherlands Antilleans, and Indonesians from the Moluccans compared with ethnic Dutch. However, South-East Asians score higher, and persons with one immigrant and one ethnic Dutch parent score only slightly below the mean of the Dutch. When comparing first-generation disadvantaged immigrant groups with later generations the data show substantial improvements for* g*, a remarkable stability of educational differences for younger children, and a clear improvement in educational achievement at the end of primary school. Indirect data on intergenerational improvements in work proficiency appear suggestive of a trend of closing gaps.*

Some of the data reflect higher cognitive capacities over time, and this enhances integration of immigrants into Dutch society. Causes of group differences and improvements in mean level of g *are discussed. Copyright © 2004 John Wiley & Sons, Ltd.*

INTRODUCTION

Group differences in *g* are rather the rule than the exception (Lynn & Vanhanen, 2002; Ogbu, 1994). In the light of these findings it is most unlikely that every group has exactly the same mean level of phenotypic cognitive capacities. Scores on tests of *g* are reflected in societal outcomes such as measures of educational achievement (Jensen, 1980; Neisser,

*Correspondence to: Jan te Nijenhuis, Open University, PO Box 2960, 6401 DL Heerlen, The Netherlands.
E-mail: JanteNijenhuis@planet.nl

Contract/grant sponsor: NWO.

Received 24 February 2003
Accepted 27 November 2003

et al., 1996) and work proficiency (Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Schmidt & Hunter, 1998); therefore, differences in *g* may account for differences between groups in these outcomes. Since the mean level of *g*, educational achievement, and work proficiency are important determinants of the speed of integration, diminishing group differences may have large consequences for groups of low-scoring non-Western immigrants. In order to find out whether this is actually taking place, a detailed description is given below of the situation in the Netherlands.

Group differences between non-Western immigrants and Dutch

There are many different immigrant groups in Europe, whereby a distinction is usually made between Western and non-Western immigrants. This paper focuses on non-Western immigrants in the Netherlands; they can be characterized by a substantial cultural distance; a lower position on the labor market and in educational settings; or both. A special group of interest is the one with mixed parents, that is one Dutch parent and one of non-Western origin; this group will be referred to as the 'mixed' group.

In recent decades growing numbers of non-Western immigrants have become part of the Dutch population, in 1999 making up 11% of the total population of about 16 million (Centraal Bureau voor de Statistiek, 2000). The largest groups come from Turkey, Morocco, the Netherlands Antilles, Surinam, and Indonesia. A special group of Indonesians is formed by second- and third-generation Moluccans, predominantly the offspring of soldiers of the Royal Dutch-Indies Army and their families who came to the Netherlands in the 1950s after the independence of the former Dutch colony, Indonesia. Immigrant children from Turkey, Morocco, the Netherlands Antilles, Surinam, and the Indonesian Moluccans islands show below average educational achievement and the unemployment rate of adult immigrants is more than double the percentage of the Dutch; only Indonesians, with the exception of the Moluccans, are not in a clearly disadvantaged position. These findings may have a lower mean level of cognitive capacities as a possible cause. This was confirmed by early studies on cognitive tests of immigrant children that yielded very low mean scores (de Jong & van Batenburg, 1984; Resing, Bleichrodt, & Drenth, 1986). The initial viewpoint of a disadvantage because of strong cultural bias in tests has now been refuted by about 30 studies (see te Nijenhuis & van der Flier, 1999, for a review of all studies up to 1999; see also Helms-Lorenz, unpublished doctoral dissertation; te Nijenhuis, 2001; van den Berg, 2001). Therefore, currently there is consensus among testing experts on the small effects of bias on the total test score, and the resulting small underestimate of *g*.

The effect of language bias

When comparing test scores of people who do not have a desirable level of proficiency in the target language and bilinguals (i.e. most immigrants) with test scores of native speakers, a distinction is usually made between verbal and nonverbal tests. Subtests with a substantial verbal component measure to an undesirable extent Dutch-language proficiency and underestimate the level of *g* of the tested non-native speakers. The more limited their proficiency in Dutch, the larger the underestimate. Using a mixture of culture-loaded and culture-reduced tests, te Nijenhuis and van der Flier (2003) found that the highly verbal subtest Vocabulary of the GATB is so strongly biased that it depresses the score on Vocabulary with 0.92 SD, leading to an underestimate of *g* based on GATB IQ by as much as 1.8 IQ points due to this single biased subtest alone, whereas the other seven subtests

combined show only very little bias and only underestimate *g* based on GATB IQ by 1.5 IQ points. So, the eight subtests of the GATB combined underestimate *g* based on GATB IQ with $1.8 + 1.5 = 3.3$ IQ points, or 0.22 SD. However, one should not forget that subtests with a strong verbal component usually constitute only a small part of a test battery; due to the use of sum scores the strong bias in tests with a verbal component becomes diluted. Moreover, looking at the effect of length of residence in the Netherlands on the scores on various intelligence tests shows the influence of language proficiency. Tests without a verbal component show small to negligible correlations with length of residence, tests with a verbal component show moderate correlations, and language proficiency tests show large correlations (see te Nijenhuis & van der Flier, 1999; van den Berg, 2001, p. 37). All these findings regarding the clear but modest role of language bias are in line with findings of language bias when testing Hispanics who do not have a desirable level of proficiency in the target language or are bilingual (Lopez, 1997; Pennock-Román, 1992).

Prediction of longer-term criteria

With concern to short-term criteria, there are no strong indications for different relationships between test scores and assessment of the quality of work and school results for immigrants and ethnic Dutch (te Nijenhuis & van der Flier, 1999; see van den Berg, 2001). How does IQ predict longer-term criteria? There are no studies into differential prediction using longer-term criteria; however, studies on over-advising using longer-term criteria are informative. When immigrant and Dutch children are matched according to their scores on school achievement or intelligence tests, the immigrant children are advised to attend higher levels of secondary education. A general finding is that several years later over-advised immigrant students have considerably higher risks of repeating classes and dropping out of school (de Jong, 1987; Driessen, 1991; Koeslag & Dronkers, 1994; Veen & Robijns, 1994). In an extensive study Tesser, Merens, and van Praag (1999) show considerable group differences in mean school marks in secondary education**,** whereby the weakest students are predominantly immigrant. When those weak students leave school, they usually end up in jobs in line with an educational level that is one level below that of the school left. It may very well be that at primary school many of the weak immigrant students in the Tesser et al. study were recommended secondary education that was one educational level higher than that advised to their Dutch classmates with a similar IQ score. This strongly suggests that the job level of over-advised immigrants is approximately at the same level as that of their Dutch classmates with comparable IQ scores in primary school. Although all of these empirical studies on over-advising are methodologically not as solid as studies on differential prediction, they suggest that for the various groups the relationships between test scores and longer-term criteria of school achievement are not greatly different.

Research question

The mean level of *g*, educational achievement, and work proficiency are important determinants of the speed of integration. Te Nijenhuis and van der Flier (2001) show in their review of the mean IQ of immigrants in the Netherlands that it is lower than the Dutch mean and that the IQs of the second-generation immigrants are rising, but do not give a detailed specification of the rise. In the meantime, several new studies have been published, including a few large-scale ones. Little research has been carried out on generational differences in educational achievement and no reviews of generational

differences in work proficiency exist. The central question addressed in this study is whether cognitive group differences are diminishing over time.

METHOD

Cook and Campbell (1979) describe models whereby research results can be generalized to other persons, settings, and times. Because random sampling for representativeness is very difficult, they opt for a series of smaller studies with haphazard samples, which they refer to as deliberate sampling for heterogeneity. The idea is to obtain heterogeneous samples of persons, settings, and times, so as to probe the robustness of a particular finding over a wide range of possible moderating factors. We give a review of Dutch studies in which data on *g*, work proficiency, and school achievement of groups of immigrants and Dutch were reported; only methodologically sound studies were included. Heterogeneous samples can vary from nationally representative to non-representative haphazard samples. Representativeness does not only refer to national representativeness, but also to local representativeness, for example when representative random samples are drawn from one city or region. Data of a complete, specific population are more interesting than data of incomplete, specific populations.

The original studies do not always report on the immigrant history, so generation sometimes had to be estimated. Adopted children were classified as first generation, because they were born outside the Netherlands. Three out of four young adult immigrants from Turkey and Morocco marry a partner from the country of origin; as these children are born in the Netherlands and grow up there, they were classified as second generation.

All methods to calculate effect sizes begin with the mean difference between two groups' test scores, and then standardize that difference by dividing it by a relevant measure of dispersion in the scores. The preferred measure of dispersion is usually the *N*-weighted average of the two groups' SDs. However, in our study the preferred way of computing effect sizes was dividing the difference between the mean score of the ethnic Dutch and the mean score of the immigrants by the SD of the Dutch. This method has two advantages: (i) this SD is not influenced by language bias, and (ii) the standardized differences derived by this method are comparable across studies, because the denominator does not fluctuate with the *N*s and SDs for the particular minority groups in the sample, as would be the case when using the *N*-weighted average of the two groups (Gottfredson, submitted). In some of the samples no SDs of the Dutch were reported, so we used the SDs of the complete group (Dutch + immigrants). As immigrant groups were usually smaller, the SD of the complete group would be quite similar to the SD of the ethnic Dutch. When the SDs of the Dutch or the complete group were not reported SDs had to be estimated in other ways; if so, this has been indicated in the tables. The resulting effect sizes of the intelligence tests were then, where appropriate, corrected for language bias, using a procedure described below.

Educational achievement

In the section on educational achievement the results of two large-scale Dutch studies are reported. Both studies used data that have been collected over several years. The first study concerns the so-called PRIMA studies (from PRIMAry school). Data from these studies were available for 1988, 1994, 1996, and 1998 and concern the educational achievement of

second- (except for 1988), fourth-, sixth-, and eighth-grade students in Dutch primary schools. The samples are nationally representative, and for each grade and year data on more than 10 000 students have been collected (Tesser & Iedema, 2001). In each grade two achievement tests were administered: in grade 2 the two tests Constructs and Arrangement, and in the other grades the two tests Language and Numerical Ability. The Constructs test appeals to the mastery of constructs such as 'many', 'few', 'before', 'behind', 'here', and 'there'. Children need these constructs not only for reading, but also for calculating. The Arrangement test contains questions that deal with classifying, seriation, comparing, and counting. Children need these skills to understand numbers and for calculation. The Language test measures achievement in reading comprehension, spelling, and vocabulary. The Numerical Ability test measures achievement in various arithmetic operations as well as insight in dealing with numbers.

The second study concerns the Primary-School Leaving Test of the National Institute for Educational Measurement (CITO). This test is administered each year to 80 000–100 000 students in the last year of primary school, which is more than half of the total population of last-year students. The Primary-School Leaving Test measures school achievement, and is used to determine the most appropriate type and level of secondary education. The test consists of the three subtests Language, Math, and Information Processing, all having the same number of items. Raw scores are transformed to standard scores with a range of 500–550 and an SD of 10. Each year new test items are constructed, and standard scores from different years are made equivalent by using anchor tests. In 1987, 1989 (Uiterwijk, 1994), and 1997 CITO collected information on the ethnicity of the students. Data on ethnicity were known for about 70% of the students in all three years; the score distributions of these subsamples were representative for the total samples. Because of the great number of Dutch students, random samples of about 5000 students were used in the analyses. The score distributions of these samples were representative of all Dutch students. For the non-Dutch students all data were included in the analyses.

Work proficiency

Studies concerning differences in work performance and job-related training results are generally based on complete, specific populations or haphazardly collected samples. An exception is the study by Dagevos (2001a) dealing with in-company promotions of employed workers. This study was based on a nationally representative sample of immigrant households in the Netherlands and a comparable (not representative) sample of majority group households. Indirect evidence for diminishing differences in work performance can be found in studies concerning unemployment and job level. These studies are typically based on large nationally representative samples using data collected by (or the under the responsibility of) the Central Bureau of Statistics (EBB 1990–2000), the Institute for Sociological–Economic Research, the Social and Cultural Plan Office (SPVA '91, '94, '98), and the Inspectorate of Work (Spijkerman, 2000). To make effect sizes for percentages (unemployed, etc.) comparable to the effect sizes for continuous variables, the standard normal curve was used to translate the percentages into z scores, from which the gap in z scores could be derived.

g

Most studies report a total IQ score, all individual subtests being weighed equally; several studies use a g score, where subtests with the highest cognitive complexity are weighed

most strongly. Additionally, data were used from studies that did not use a whole test battery, but measured only one or a few narrow mental abilities.

When comparing group scores on various tests, one should be aware of several psychometric traps (Jensen, 1980). Cognitive tests differ in their *g* loadings; when groups differ in their mean level of *g*, it is therefore unlikely that the group differences on all subtests of a battery are equally large (Jensen, 1985, 1998). So, when interpreting group differences these differences in *g* loadings of subtest have to be taken into account. The total score of a classic IQ battery correlates about 0.95 with the *g* score obtained from the same battery (Jensen, 1998, p. 90). The individual subtests of an IQ battery correlate substantially lower than 0.95 with the total score; their *g* loadedness is therefore lower than 0.95. Of the tests that measure narrow mental abilities only Raven's Progressive Matrices has an exceptionally high *g* loading. So, tests of narrow mental abilities and subtests of a battery reflect *g* only partially; therefore, one would expect smaller group differences than on a test battery. Based on the empirical work of Carroll (1993), the *g* loadings of tests of narrow mental abilities will be estimated, and used to estimate group differences in *g*. Carroll (1993, p. 597, Table 15.5) reports the average loadings of first-order factors on third-order *g* factors. These values will be taken as estimates of the *g* loadings of the tests in our review. Subsequently, the mean difference in test score between immigrants and native Dutch, expressed in SDs, will be multiplied by the factor 1/*g*, resulting in an estimate of the group difference in *g*. Tests with a substantial verbal component measure both *g* and Dutch language proficiency for many immigrants, so the correction method described above needs to be combined with a correction for language bias, which is described below. Apart from *g*, tests also measure one or more second-order factors, which may reflect group differences. These will influence to a certain extent the estimates in SDs. The *g* loading of a composite of two tests was computed using the formula given by Jensen (1998, p. 104).

As mentioned above, mean IQ scores may be underestimates of *g* as a result of language bias. To our knowledge, there is no procedure for a language bias correction, so we made an attempt to develop such a procedure. We made estimations of the biasing effects, based on two studies from te Nijenhuis and van der Flier (1997, 2003), who used GATB data on first-generation immigrants. Te Nijenhuis and van der Flier (2003) estimate the biasing effects for Turks and Moroccans to be about three IQ points; no estimates were given for Surinamese or Antilleans. However, from te Nijenhuis and van der Flier (1997) a measure of bias can be constructed by looking at the differences in scores between two subtests with equal *g* loadings, namely Vocabulary and Arithmetic Reason, the first of which is sensitive to language bias. For Surinamese the scores on the two tests are practically the same, and the difference for Antilleans (0.27 SD) is about one-third of the difference for Turks (0.72 SD). This confirms the absence of language bias for Surinamese and the strong bias for Turks, and suggests a substantial bias for Antilleans.

In order to assign corrections to each data point a few decisions had to be made. Concerning first-generation immigrants, we decided not to correct for bias with samples of Surinamese and to apply a correction for Antilleans of one IQ point (0.07 SD), and for Turks and Moroccans of three IQ points (0.20 SD) for test batteries that resemble broad, classical test batteries such as the GATB. No corrections were applied to nonverbal tests batteries or nonverbal single tests. When only one test with a substantial verbal component was used, a correction of 0.92 SD was applied for Turks and Moroccans, the estimated amount of language bias in the GATB subtest Vocabulary (te Nijenhuis & van der Flier,

2003) and a correction of 0.92 SD/3 for Antilleans (see above). With concern to samples of second-generation immigrants, corrections were only applied for young children with Turkish and Moroccan parents, who tend to grow up in a non-Dutch-speaking environment; Helms-Lorenz (unpublished doctoral dissertation, ch. 3) reports amounts of test bias for this group that are comparable to the bias reported for first-generation young children. For samples that consisted of a mix of first- and second-generation immigrants, a value was taken that was between the value for the first generation and the value for the second generation.

As there are no data on test bias for groups of Indonesians from the Moluccans and South-East Asians, we made no corrections for these groups. How to deal with samples that combine persons from various groups? When the four largest disadvantaged immigrant groups are combined into one group it is made up of 30% Turks, 25% Moroccans, 34% Surinamese, and 11% Antilleans (Centraal Bureau voor de Statistiek, 1996). So, we reasoned that samples of Surinamese and Antilleans would be most likely made up of 25% of Antilleans, leading to a correction of 0.02 SD for samples of first-generation Surinamese and Antilleans. For the samples classified as various, the value of the overall correction for bias was computed by multiplying the percentage of the disadvantaged immigrant population for a group with the value of the correction for that specific group and then summing the products. For instance, for a group of various young children of the first generation, the correction for bias was computed as 55% of 0.20 SD for the young Turks and Moroccans + 34% of 0.00 SD for the Surinamese + 11% of 0.07 SD for the Antilleans = 0.12 SD. For a sample of both young and older children it was assumed that 50% were young children and 50% were older children. Again, for samples that consisted of a mix of first- and second-generation immigrants, a value was taken that was between the value for the first generation and the value for the second generation. The study of Verhaeghe et al. (2002) was based on both Dutch and Belgian children, and as no Belgian bias research has been published we decided to base the corrections for language bias on the Dutch findings. The Verbal Reasoning test of the Verouden, Ross, Stet, and Scheele (1987) study was corrected using the formula 55% of 0.92 SD (T.+M.) + 34% of 0.00 SD (S.) + 11% of 0.92 SD/3 (Ant.), leading to a value of 0.54 SD for the correction. Even more fine-grained corrections for bias might have been applied, but considering that the effect sizes of the corrections were mostly small or very small, more fine-grained corrections would differ by only a few hundredths of an SD from the present corrections, so we deemed them unnecessary. So, for all studies we report an empirically established effect size and for a number of studies with samples that empirically have shown sensivitivy to language bias or samples that are comparable to them we also report an effect size that is corrected for language bias; together they put the effect sizes of groups with and without a desirable level of proficiency in the target language on a common metric. Table 1 gives a review of the most common corrections.

RESULTS

Results are given for educational achievement, work proficiency, and *g*. Tables 2–4 list the studies in chronological order, but the text has its own dynamics, so the descriptions of the studies in the text are not always chronological.

Table 1. Corrections for bias by group and generation expressed in SDs

Group	Generation		
	1	1/2	2
Turks and Moroccans			
Young children	0.20	0.20	0.20
Older children	0.20	0.10	
Adults	0.20	0.10	
Surinamese			
Antilleans	0.07	0.04	
Surinamese and Antilleans	0.02	0.01	
Various			
Young children	0.12	0.12	0.11
Older children	0.12	0.06	
Young and older children	0.12	0.09	0.06
Adults	0.12	0.06	

Generation 1/2 refers to immigrants from both the 1st and the 2nd generation.

Educational achievement

The first part of Table 2 shows the data of the PRIMA studies. The effect sizes for the verbally loaded subtests Constructs and Language are in the ranges of, respectively, 1.35–1.65 for Turks (median value 1.49); 1.06–1.61 for Moroccans (median value 1.19); 0.55–0.92 for Surinamese (median value 0.67); 0.85–1.22 for Antilleans (median value 0.98); and 0.15–0.37 for the mixed group (median value 0.24). The effect sizes for the subtests Arrangement and Numerical Ability are in the ranges of, respectively, 0.55–0.99 for Turks (median value 0.80); 0.64–1.09 for Moroccans (median value 0.82); 0.54–0.85 for Surinamese (median value 0.63); 0.63–1.15 for Antilleans (median value 0.94); and 0.15–0.38 for the mixed group (median value 0.23). So, on all measures, all immigrant groups score lower than the Dutch. Turks, Moroccans, Surinamese, and Antilleans score substantially lower that the Dutch, but the mix group scores only a quarter of an SD lower. For the immigrant students from the four largest groups Surinamese systematically outscore Turks, Moroccans, and Antilleans on all tests. On the language tests Antilleans outscore Turks and Moroccans, but on the numerical tests Turks and Moroccans generally outscore Antilleans. Looking at the profile of verbal test scores and numerical test scores the Surinamese and Antillean groups have roughly the same scores in comparison with the mean of the Dutch. However, for the Turkish and Moroccan groups the scores on the verbal tests are much lower than the scores on the numerical tests, indicating a strong influence of proficiency in Dutch. Over a decade, the differences have been remarkably consistent for the early grades, but have become smaller for grade 8: the scores improve with about a third of a standard deviation for 11–12 year olds. The scores of the mixed group have remained quite steadily one quarter of a standard deviation below the mean of the Dutch.

The second part of Table 2 shows the data on the Primary-School Leaving Test of the National Institute for Educational Measurement (CITO) for 11–12 year olds. As already reported by de Jong (1987) these immigrant children tend to have levels of educational achievement that resemble those of ethnic Dutch children that are 1 or 2 years younger. The Surinamese/Antillean group shows the smallest differences. The differences in mean scores between Dutch and immigrant students in 1987 and 1989 are the same. In 1997 these differences had become about 0.15 SD smaller. The effect

Table 2. Differences in mean scores on school achievement tests between immigrants and Dutch in Dutch studies

Study/subtest/grade	Group	*n*				Effect size				ES_{med}
Tesser and Iedema (2001)										
Constructs										
Second grade			1994	1996	1998		1994	1996	1998	
(5–6 years old)	Turks		1279	1507	1241		1.46	1.46	1.41	1.44
	Moroccans		999	1257	1157		1.20	1.11	1.22	1.18
	Surinamese		364	514	494		0.66	0.67	0.70	0.68
	Antilleans		70	124	170		0.93	1.06	1.12	1.04
	Mix		582	791	380		0.23	0.26	0.30	0.26
Language										
Fourth grade		1988	1994	1996	1998	1988	1994	1996	1998	
(7–8 years old)	Turks	820	1256	1263	1256	1.51	1.54	1.65	1.60	1.58
	Moroccans	832	1092	1224	1059	1.23	1.12	1.19	1.22	1.19
	Surinamese	506	650	522	573	0.66	0.55	0.67	0.59	0.62
	Antilleans	77	90	93	151	0.85	0.95	0.98	0.89	0.92
	Mix	924	697	670	329	0.34	0.19	0.19	0.15	0.22
Language										
Sixth grade		1988	1994	1996	1998	1988	1994	1996	1998	
(9–10 years old)	Turks	759	759	1086	1029	1.48	1.42	1.49	1.54	1.48
	Moroccans	791	852	1088	1016	1.18	1.18	1.09	1.10	1.14
	Surinamese	505	585	571	591	0.63	0.66	0.64	0.71	0.66
	Antilleans	73	80	67	101	0.91	0.85	1.12	0.93	0.95
	Mix	920	822	651	295	0.29	0.24	0.19	0.26	0.25
Language										
Eighth grade		1988	1994	1996	1998	1988	1994	1996	1998	
(11–12 years old)	Turks	689	947	967	906	1.65	1.49	1.35	1.39	1.47
	Moroccans	723	843	902	857	1.61	1.25	1.20	1.06	1.28
	Surinamese	558	599	530	587	0.92	0.92	0.75	0.70	0.82
	Antilleans	59	71	82	102	1.22	1.04	1.08	1.11	1.11
	Mix	757	732	712	266	0.24	0.37	0.18	0.16	0.24

Continues

Table 2. Continued

Study/subtest/grade	Group	*n*				Effect size				ES_{med}
Arrangement										
Second grade			1994	1996	1998	1988	1994	1996	1998	
(5–6 years old)	Turks		1272	1501	1241		0.99	0.97	0.90	0.95
	Moroccans		996	1243	1143		0.91	0.84	0.96	0.90
	Surinamese		366	496	512		0.60	0.58	0.56	0.58
	Antilleans		74	120	172		0.63	0.77	0.87	0.76
	Mix		599	775	388		0.23	0.19	0.23	0.21
Numerical ability										
Fourth grade		1988	1994	1996	1998	1988	1994	1996	1998	
(7–8 years old)	Turks	805	1261	1268	1208	0.95	0.70	0.69	0.80	0.79
	Moroccans	823	1108	1215	1004	0.86	0.64	0.71	0.86	0.77
	Surinamese	496	655	523	542	0.65	0.61	0.63	0.70	0.65
	Antilleans	75	88	94	147	1.04	0.94	1.07	0.88	0.98
	Mix	895	681	682	322	0.38	0.15	0.24	0.19	0.24
Numerical ability										
Sixth grade		1988	1994	1996	1998	1988	1994	1996	1998	
(9–10 years old)	Turks	752	1047	1086	939	0.98	0.64	0.64	0.65	0.73
	Moroccans	800	852	1089	905	1.03	0.69	0.65	0.80	0.79
	Surinamese	502	585	565	523	0.73	0.63	0.54	0.71	0.65
	Antilleans	74	76	65	89	1.15	0.77	1.09	0.98	1.00
	Mix	913	830	642	273	0.32	0.21	0.15	0.31	0.25
Numerical ability										
Eight grade		1988	1994	1996	1998	1988	1994	1996	1998	
(11–12 years old)	Turks	685	959	964	850	0.98	0.80	0.60	0.55	0.73
	Moroccans	722	840	906	761	1.09	0.82	0.75	0.65	0.83
	Surinamese	556	597	520	534	0.85	0.81	0.66	0.59	0.73
	Antilleans	61	69	82	81	0.97	0.92	1.15	0.82	0.97
	Mix	747	730	711	249	0.27	0.30	0.18	0.19	0.24

Uiterwijk (1994) (1997 data not published)								
		1987	1989	1997	1987	1989	1997	
Eighth grade	Turks	797	919	220	1.16	1.25	0.95	1.13
(11–12 years old)	Moroccans	720	907	148	1.18	1.21	1.08	1.16
	Surinamese/Antilleans	829	832	165	0.96	0.98	0.83	0.92

The data refer to immigrants from both the 1st and the 2nd generation.
*N*s refer to the immigrant groups.
Effect sizes: Dutch mean minus immigrant group mean, and divided by SD of total reference sample; ES_{med} = median effect size. These values are not corrected for language bias.

Table 3. Differences in mean job proficiency between Dutch and immigrants in Dutch studies

Study	Group	*n*	Job/criterion	Generation	Effect size
van Rooijen (1992)	Surinamese		*Tram drivers*		
		18	Performance ratings	1	0.42
		15	Accidents	1	+0.18
		17	Absenteeism	1	0.05
van Gastel, Bleichrodt,	Various		*ITC job trainees*		
and van Heelsum (1994)		63	% employed (after 5 years)	1/2	0.18
		63	No. of subordinates	1/2	+0.07
van Leest (1997)	Turks/		*Police officer trainees*		
	Moroccans	107	Progress in training	1/2	0.50
de Vries and	Various		*Police officers*		
Pettigrew (1998)		29	Grades	1	0.40
		24	Performance ratings	1	0.27
		27	Writing reports	1	0.88
			Nurses		
		32	Grades	1	0.40
		28	Performance ratings	1	0.49
te Nijenhuis and	Various		*Trainee truck drivers*		
van der Flier (2000)		78	Examination results	1	0.27
		78	Professional attitude	1	0.40
		78	Practice skills	1	0.19
van den Berg (2001)	Various		*Employment agency trainees*		
		234	General capacities	1/2	0.48
		234	Interpersonal skills	1/2	0.23
		234	Work attitude	1/2	0.19
		234	Work performance	1/2	0.21
		234	Language skills	1/2	0.69
		234	Total score	1/2	0.43
	Turks/ Moroccans	25	Total score	1/2	0.31
	Surinamese/ Antilleans	59	Total score	1/2	0.22
	Various	124	Total score	1	0.31
	Various	21	Total score	2	0.26
	Various	234	Absenteeism	1/2	0.04
Dagevos (2001a)			*Employed workers*		
	Turks	1123	% internal promotion	1/2	0.48
	Moroccans	866	% internal promotion	1/2	0.21
	Surinamese	1651	% internal promotion	1/2	0.13
	Antilleans	746	% internal promotion	1/2	0.24

*N*s refer to the immigrant group sample sizes. Effect size: majority group mean minus immigrant group mean, divided by majority group SD. For percentages, difference between corresponding *z* scores under the normal curve. Assignment to first or second generation based on description in original study. When no clear descriptions were available, generation was estimated. Generation 1/2 means testees from both the 1st and the 2nd generation. Higher scores for immigrants have the prefix '+'; lower scores for immigrants have no prefix.

size for the difference between Dutch and immigrant groups in 1997 still was about one standard deviation.

Work proficiency

Studies concerning differences in work performance and job-related training results are summarized in Table 3. Van Rooijen (1992) compared the job performance of all new

immigrant streetcar drivers in Amsterdam in the period 1977–1986 ($n = 20$) and a random sample of all their colleagues of the majority group ($n = 132$). Mean performance ratings of the immigrants were 0.42 SD lower. For accidents and absenteeism only small differences were found between the two groups. Van Gastel et al. (1994) compared the careers of highly educated immigrants ($n = 63$) and majority group members ($n = 87$) who had participated in a job counseling/vocational training program for the unemployed five years earlier. Only small differences in career success were found. A study by van Leest (1997) into the selection of police officers revealed that immigrants ($n = 107$) had less positive training results than majority group members ($n = 260$), the effect size being 0.50 SD. De Vries and Pettigrew (1998) collected grades and performance ratings of matched groups of immigrant and majority group police officers ($n = 31$ for both groups) and nurses ($n = 32$ for both groups). Effect sizes for grades and performance ratings centred around 0.40 with a relatively high value of 0.88 for writing reports in the group of police officers. Te Nijenhuis and van der Flier (2000) present results of a study among all the immigrant trainee truck drivers ($n = 78$) in an organization with offices throughout the country, matched with the same number of trainee truck drivers of the majority group. Effect sizes vary from 0.19 SD for practice skills to 0.27 for examination results and 0.40 for professional attitude. Van den Berg (2001) studied training performance of participants in a job-specific vocational-training program organized by one of the regional Dutch employment agencies. The sample consisted of 234 immigrants and 213 majority group members. Based on information about their background a distinction was made within the group of immigrants between Turks/Moroccans and Surinamese/Antilleans, and also between first- and second-generation immigrants. The immigrant group was rated substantially lower than the majority group. The largest difference was found for language proficiency (0.69 SD). For the total score on a behavioral observation scale covering the various performance aspects the difference was 0.43 SD. Differences in total scores were larger for Turks/Moroccans than for Surinamese/Antilleans (effect sizes of 0.31 and 0.22, respectively) and larger for first-generation immigrants than for second-generation immigrants (effect sizes of 0.31 and 0.26, respectively).

In a study by Dagevos (2001b) based on a large representative sample of immigrant households in the Netherlands and a comparable (not-representative) sample of majority group households a comparison was made between the percentages of internal promotions of employed workers. Percentages were lower for the immigrant groups. Effect sizes vary from 0.13 for Surinamese to 0.48 for Turks.

There are no data available that allow strong conclusions with respect to diminishing group differences over time in work performance between immigrants and majority group members. Indirect evidence can, however, be found in studies concerning unemployment and job level. The results are summarized in Table 4. Longitudinal comparisons over time will be discussed first, followed by comparisons between first- and second-generation immigrants.

Based on very large samples from the Central Bureau of Statistics (EBB 1990–2000) Dagevos (2001b) gives an overview of unemployment rates and percentages of workers with low-level jobs in the various immigrant groups in the period of 1990–1999/2000. Differences in unemployment rates between immigrants and majority group members are strongly diminishing. Effect sizes in 2000 appear to be 30–50% lower than in 1990. Differences in percentages of workers with low-level jobs are also diminishing; however, the decrease in effect sizes is not very strong. Two studies by Dagevos (1996, 2001b), based on data from representative samples of households, allow us to compare differences

Table 4. Changes in differences between Dutch and immigrants on job proficiency-related variables in Dutch studies

Study	Group	*n*	Variable	Generation	Effect size
Dagevos (2001b)	Turks	839/726	% unemployed 1990/2000	1/2	1.01/0.54
	Moroccans	396/339	% unemployed 1990/2000	1/2	1.12/0.75
	Surinamese	740/731	% unemployed 1990/2000	1/2	0.82/0.54
	Antilleans	230/289	% unemployed 1990/2000	1/2	0.72/0.47
	Turks	564/461	% low level jobs 1990/1999	1/2	1.04/0.88
	Moroccans	302/279*	% low level jobs 1990/1999	1/2	1.00/0.89
	Surinamese	532/460	% low level jobs 1990/1999	1/2	0.49/0.42
	Antilleans	176/159	% low level jobs 1990/1999	1/2	0.24/0.32
Dagevos (1996)/Dagevos (2001b)	Turks	574/1613	Mean job level 1991/1998	1/2	1.09/0.96
	Moroccans	390/1473	Mean job level 1991/1998	1/2	0.95/1.03
	Surinamese	348/1856	Mean job level 1991/1998	1/2	0.47/0.44
	Antilleans	236/1079	Mean job level 1991/1998	1/2	0.26/0.37
Spijkerman (2000)	Various	2150	Mean job level 1998	1/2	0.79
Dagevos and Veenman (1996)	Turks	662/103	% unemployed 1st/2nd gen.	1 or 2	0.83/0.72
	Moroccans	564/71	% unemployed 1st/2nd gen.	1 or 2	0.94/0.92
	Surinamese	645/175	% unemployed 1st/2nd gen.	1 or 2	0.59/0.50
	Antilleans	410/145	% unemployed 1st/2nd gen.	1 or 2	0.78/0.07
Tesser et al. (1999)†	Turks	438/213	Mean job level 1st/2nd gen.	1 or 2	1.07/0.70
	Moroccans	438/123	Mean job level 1st/2nd gen.	1 or 2	1.20/0.80
	Surinamese	772/313	Mean job level 1st/2nd gen.	1 or 2	0.49/0.35
	Antilleans	423/145	Mean job level 1st/2nd gen.	1 or 2	0.53/0.02

*N*s refer to the the immigrant group sample sizes. Effect size: majority group mean minus immigrant group mean, divided by majority group SD. For percentages, difference between corresponding *z* scores under the normal curve. Effect sizes to the left and the right of the / symbol correspond to the labels to the left and the right of the / symbol in the variable column. Assignment to first or second generation based on description in original study. When no clear descriptions were available, generation was estimated. Generation 1/2 means testees from both the 1st and the 2nd generation. Generation 1 or 2 means testees from the 1st and the 2nd generation, respectively.

*Estimated sample sizes.

†Published in Tesser et al. (1999), but not exactly in this way. Tesser et al. differentiated between first generation immigrants who immigrated before 1981 and immigrants since 1981. To combine the results of these subgroups we had to go back to the original SPVA 1998 data.

in mean job level between immigrants and majority group members in 1991 and 1998. Inspection of effect sizes does not reveal an overall trend in the direction of smaller differences in 1998. An independent study by Spijkerman (2000), based on a representative national sample of the work force ($n = 41\,083$) and carried out by the Inspectorate of Work, shows an effect size for the difference in job level between immigrants and majority group members in 1998 that is comparable to the effect size in the Dagevos studies.

Information about changes in unemployment rate and mean job level between first- and second-generation immigrants is provided in studies by Dagevos and Veenman (1996) and Tesser et al. (1999). Both studies are based on nationally representative samples of households. Unemployment rates for second-generation immigrants are closer to those of the majority group. This effect is strongest for the Antilleans (0.78 and 0.07, respectively, for the first and the second generation). With regard to mean job level there is a clear improvement for all second-generation immigrants. The decrease in effect sizes in comparison with first-generation immigrants is about one-third of a standard deviation. The mean job level of second-generation Antilleans is even similar to that of the majority group.

g

Nationally and locally representative samples

With concern to test batteries, there are three studies that use samples that are representative of the Dutch population, namely Resing et al. (1986; reanalysed by te Nijenhuis, Tolboom, Resing, and Bleichrodt, in press); Snijders, Tellegen, and Laros (1988); and Tellegen, Winkel, Wijnberg-Williams, and Laros (1998). Snijders et al. and Tellegen et al. used representative sampling procedures for their immigrant samples, but the sizes of their immigrant groups are rather small; therefore, small sampling errors may have large consequences. Because of its careful sampling and its large numbers the immigrant sample of Resing et al./te Nijenhuis et al. is the best approximation of a representative sample of immigrant children.

In a fourth study, on the WISC III, by Verhaeghe et al. (2002), an attempt was made to use a sample that was representative for the combined Dutch and Flemish population. The information Verhaeghe et al. supply is limited, but suggests an attempt was made to have a nationally representative sample of immigrants as well. The sampling procedure employed appears to have been suboptimal, and may have led to a small underestimate of *g* (Tellegen, 2002). As there are no signs that the flaws in the sampling procedure affect the immigrants and the majority group differently and as we use a difference score in the present study, the Verhaeghe et al. study yields useful information.

With concern to tests for narrow mental abilities, Tesser et al. (1999) used data from a national cohort study on primary education. Due to the careful sampling and the large numbers ($N = 49\,085$), it is a good approximation of a nationally representative sample for both ethnic Dutch and immigrant children. Children of, respectively, 8, 10, and 12 years took a nonverbal cognitive test, consisting of two parts: a test of Spatial Relations (g_v) with an estimated *g* loading of 0.40, and a test of Induction (g_{fl}) with an estimated *g* loading of 0.57. Assuming that the sum score measures *g* more reliably than the two individual subtests (Nunnally, 1978; Lubinski & Dawis, 1992), we computed the *g* loading of the sum score using the formula reported by Jensen (1998, pp. 103/104), leading to a value of 0.63.

Local representativeness was achieved in the study of de Jong and van Batenburg (1984), in which random representative samples were drawn from all primary schools in the city of Rotterdam.

For these six carefully designed studies of representative samples mentioned above Table 5 shows that for the first-generation sample the effect sizes range from, respectively, 0.93 to 1.50 for Turks and Moroccans (weighted average 1.28), and 0.93 to 1.07 for Surinamese and Antilleans (weighted average 1.00). South-East Asians' ES is +0.62 SD, meaning they score higher. For the combination of first- and second-generation samples, the effect sizes range from, respectively, 0.60 to 1.16 for Turks and Moroccans (weighted average 0.85), and 0.21 to 0.70 for Surinamese and Antilleans (weighted average 0.51). For the second-generation samples, the effect sizes range from, respectively, 0.69 to 1.23 for Turks and Moroccans (weighted average 0.83), and 0.75 to 0.89 for Surinamese and Antilleans (weighted average 0.77). The ES is 0.63 for the various group and 0.08 for the mixed group. For the second- and third-generation sample of Moluccans, the effect size is 0.40.

A comparison of the weighted averages from the first and the second generation reveals that the group differences are clearly diminishing over time for Turks and Moroccans and for the Surinamese and Antilleans. The means of the samples that include both first- and second-generation immigrants appear to contradict this finding, because these means are not in between the means of the first- and second-generation samples used in the other studies. However, the means of the samples that include both first- and second-generation immigrants are completely based on small samples using the nonverbal SON-R. Of the six nationally and locally representative studies only Resing et al./te Nijenhuis et al. has sample sizes that are quite large and includes both first- and second-generation children, making it the study most convincingly showing diminishing group differences. So, it is concluded, based on the findings from six nationally and locally representative samples, that immigrants' *g* increases over time.

Complete, specific populations

With concern to test batteries, te Nijenhuis and van der Flier (1997) compared the test results of all the immigrant job applicants of the Dutch Railways and regional bus companies between 1988 and 1992 with those of a random representative sample of all Dutch applicants over the same period of time. Te Nijenhuis and van der Flier (2000) present results of a study among all the immigrant trainee truck drivers in an organization with offices throughout the country, matched with the same number of Dutch trainee truck drivers. Van Rooijen (1992) used the data of all immigrant streetcar drivers in Amsterdam and a random sample of all their ethnic Dutch colleagues. Snijders et al. (1988) tested the entire Dutch population of deaf children within a specific age range. It is remarkable that of this population 18% were immigrant children.

Table 5 shows that for the first-generation sample the effect sizes range from, respectively, 1.23 to 1.66 for Turks and Moroccans (weighted average 1.39), and 1.05 to 1.10 for Surinamese and Antilleans (weighted average 1.08). The effect size is 0.87 for the various group. For the combination of first- and second-generation samples, the effect size is 0.89 for Turks and Moroccans and 0.67 for Surinamese and Antilleans.

The effect sizes are much smaller for the sample including second-generation immigrants. However, of the studies using complete, specific populations all first-generation scores are on the broad, classical GATB, and all 1/2-generation scores are on the nonverbal SON-R, thereby possibly exaggerating the diminishment of scores over time. Still, when taking the difference between broad and nonverbal batteries into account, there is a clear improvement in *g* for the second generation. So, the group differences are diminishing over time. These findings are in line with the findings from studies using representative samples.

Table 5. Differences in mean scores on intelligence test batteries and tests of special intelligence factors between immigrants and Dutch in Dutch studies

Study	Sample	Group	*n*	Test	Test type	Gen.	ES	ES_{adj}
	Nationally and locally representative samples							
de Jong and van	Young children	Turks and Mor.	106	GALO	Broad battery	1	1.13	0.93
Batenburg (1984)		Surinamese	110			1	0.93	
Resing et al.	Young children	Turks	130	RAKIT	Broad battery	1	1.45	1.25
(1986*)/te		Moroccans	177			1	1.70	1.50
Nijenhuis et al.		Sur. and Ant.	123			1	1.09	1.07
(in press)		Turks	104			2	1.20	1.00
		Moroccans	76			2	1.43	1.23
		Sur. and Ant.	71			2	0.77	
Snijders et al. (1988)†	Young and older children	Turks	24	SON-R	Nonverbal battery	1/2	1.16	
	Older children	Moroccans	9	5.5-17		1/2	0.82	
		Sur. and Ant.	10			1/2	0.22	
Tellegen et al. (1998)‡	Young children	Turks	22	SON-R	Nonverbal battery	1/2	0.60	
		Moroccans	25	2.5-7		1/2	0.79	
		Surinamese	36			1/2	0.70	
		Antilleans	15			1/2	0.21	
Tesser et al. (1999)§	Young children	Turks	2872	Spatial	2 nonverbal tests	2	0.69	
		Moroccans	2411	Relations +		2	0.97	
		Surinamese	1393	Induction		2	0.75	
		Antilleans	232			2	0.89	
		Moluccans	84			2/3	0.40	
		S.-E. Asians	267			1	+0.62	
		Mix	1999			2	0.08	
Verhaeghe et al. (2002)‖	Young and old. children	Various	128	WISC III	Broad battery	2	0.70	0.64
	Complete, specific populations							
Snijders et al. (1988)†	Hearing-impaired	Turks and Mor.	31	SON-R	Nonverbal battery	1/2	0.89	
	Young and older children	Sur. and Ant.	85	5.5-17		1/2	0.67	
van Rooijen (1992)	Streetcar drivers	Surinamese	20	GATB	Broad battery	1	1.05	
te Nijenhuis and van der	Applicants at	Turks	275	GATB	Broad battery	1	1.43	1.23

Continues

Table 5. Continued

Study	Sample	Group	*n*	Test	Test type	Gen.	ES	ES_{adj}
Flier (1997)¶	Dutch railways	Moroccans	167			1	1.86	1.66
		Surinamese	535			1	1.08	
		Antilleans	126			1	1.17	1.10
te Nijenhuis and van der Flier (2000)	Trainee truck drivers	Various	78	GATB	Broad battery	1	0.99	0.87
	Haphazardly collected samples							
Verouden et al. (1987)	Applicant ticket collect.	Various	364	Verbal Reas.	Verbal test	1	2.02	1.48
van de Vijver, Willemse, and van de Rijt (1993)	Young children	Various	32	WISC-R	Broad battery	1	1.74	1.62
			26			2	0.81	0.70
		Mix	21			2	0.17	
Hessels (unpublished doctoral dissertation)	Young children	Turks	198	RAKIT	Broad battery	1/2	1.47	1.27
		Moroccans	199	(short form)		1/2	1.38	1.18
van Leest and Dutman (1995)**	Trainee police officers	Turks and Mor.	65	PIT	Broad battery	1	1.00	0.80
te Nijenhuis, Evers, and Mur (2000)	Older children	Various	111	DAT	Broad battery	1	1.14	1.02
van de Vijfeijken and Vedder (2000)††	Young children	Turks	42	Draw-A-Man	Nonverbal test	2	+0.58	
		Moroccans	14			2	+0.16	
Stams, Juffers, Rispens, and Hoksbergen (2000)	Young, adopted children	Sri Lankans	100	RAKIT	Broad battery	1	+0.27	
		Colombians	11	(short form)		1	+0.80	
		Koreans	36			1	+1.00	
Helms-Lorenz (unpublished doctoral dissertation)	Young children	Turks	4	RAKIT (short form)	Broad battery	1	1.82	1.62
Helms-Lorenz et al. (2003)‡‡		Moroccans	6			1	0.84	0.64
		Sur. and Ant.	7			1	0.76	0.74
		Mix	6			1	0.17	
		Turks	7	SON-R	Nonverbal battery	1	0.64	
		Moroccans	27	(short form)		1	1.40	
		Sur. and Ant.	17			1	1.49	
		Mix	12			1	0.35	

		Turks	34	RAKIT	Broad battery	2	1.43	1.23
		Moroccans	29	(short form)		2	0.98	0.78
		Sur. and Ant.	11			2	0.64	
		Mix	10			2	0.86	
		Turks	31	SON-R	Nonverbal battery	2	0.52	
		Moroccans	65	(short form)		2	0.65	
		Sur. and Ant.	11			2	0.46	
		Mix	11			2	+0.13	
van den Berg (2001)/van	Applicants for	Various	2254	MCT-M	Broad battery	1	1.20	1.08
den Berg and Bleichrodt	government job	Various	288			2	0.38	
(2000)[§§]	training programs	Mix	152			2	+0.01	

Ns refer to the immigrant group sample sizes.

With concern to test type, a broad test battery is a classical test battery that measures several second-order abilities and includes at least one verbal subtest.

Effect sizes: Dutch mean minus immigrant group mean, and divided by Dutch SD. Effect sizes for tests of one or two narrow mental abilities: extrapolation to $g = 1.00$. For procedure of estimating *g* loadings and group differences in *g*, see text. Estimated *g* loadings of tests, respectively, combination of Spatial Relations and Induction tests (Tesser et al., 1999), $g = 0.63$; Verbal Reasoning measures Induction, $g = 0.57$ (Verouden, Ross, Stet, & Scheele, 1987). ES_{adj} = effect size adjusted for language bias. See text for details. Higher scores for immigrants have the prefix '+', lower scores for immigrants have no prefix.

Assignment to first or second generation based on description in original study. When no clear descriptions were available, generation was estimated. See text for details. Generation 1/2 means testees from both the 1st and the 2nd generation.

*Dataset from Resing et al. (1986) reanalysed by te Nijenhuis et al. (in press).

†Snijders et al. (1988): data on deaf children; differences in mean between deaf majority group children and deaf immigrant children.

‡The sample of Tellegen et al. (1998) consisted of 76 per cent second-generation children.

§Tesser et al. (1999): numbers within each age group were highly comparable, so the effect sizes were averaged over the three age groups.

‖WISC-III data are based on a combination of a Dutch and a Flemish sample. The complete sample consisted of 1600 children, of which eight per cent was immigrant. Figures are based on the press release from Ghent University (Verhaeghe et al., 2002) and personal communication of Willem Kort with the first author.

¶Effect sizes are not reported in the te Nijenhuis and van der Flier (1997) study, but were computed by the first author.

**Van Leest and Dutman (1995) used the Police Intelligence Test, a measure of general intelligence. Group differences are not reported, but were described by van Leest as about a standard deviation (personal communication with the first author).

††Means and SDs were not reported in the article, but were supplied by van de Vijfeijken to the first author.

‡‡Helms-Lorenz (unpublished doctoral dissertation) reports only her data on second-generation immigrant children, but supplied us with the full dataset, which also includes first-generation immigrant children.

§§In their study with the MCT-M van den Berg (2001) and van den Berg and Bleichrodt (2001) did not report means and did not split up their immigrant sample into generations, but these data were supplied by the authors. Additional, more recent data were added to the data used in the two published studies.

Haphazardly collected samples

With concern to test batteries, the remaining studies (Helms-Lorenz, unpublished doctoral dissertation/Helms-Lorenz et al., 2003; Hessels, unpublished doctoral dissertation; Stams et al., 2000; van de Vijver et al., 1993; van den Berg & Bleichrodt, 2000/van den Berg, 2001; van Leest & Dutman, 1995) clearly used haphazard sampling. Helms-Lorenz, Hessels, and van de Vijver et al. used samples of primary-school children that were at hand. Though in Helms-Lorenz (unpublished doctoral dissertation) only data on second-generation immigrant children are reported, the dataset also includes first-generation children. In this paper we report on both. The sample by van den Berg and Bleichrodt of job applicants resembles that of the previous section on complete, specific samples, the difference being that the specific population was not completely sampled. These authors supplied additional, more recent data, that were added to the data used in the two published studies. An interesting point in this large dataset is that there are data on the age first-generation immigrants came to the Netherlands. Those who came before the age of 7 ($n = 168$) scored 0.49 SD lower and those that came after the age of 7 ($n = 757$) scored 1.20 SD lower than the Dutch. So, the younger the age of entry into the Netherlands, the higher the IQ score. Van Leest and Dutman (1995) studied immigrant trainee police officers which were not directly compared with ethnic Dutch in the same training program. Group differences are not reported, but were described by van Leest as approximately one standard deviation (personal communication with the first author). Stams et al. (2000) studied 7-year-old non-European adopted children.

With concern to tests for narrow mental abilities, van de Vijfeijken and Vedder (2000) studied the Draw-A-Man test with 7- and 8-year-old children. The immigrant children in the van de Vijfeijken and Vedder study scored 0.16–0.58 SD higher than the Dutch children. This is the only study in which Turkish and Moroccan ($N = 56$) children outscore Dutch children ($N = 53$). However, in a study performed in the Caribbean using the same test van de Vijfeijken, Vedder, and Kook (1997) found that Netherlands Antillean children scored lower than Dutch children. Harrison and Schock (1994) argue that the test is not a good measure of intelligence, and according to Jensen (1980) the test cannot be used as a culture-fair test, because it is sensitive to cultural influences. We therefore left these findings out when estimating effect sizes per generation.

In an unpublished report, Verouden et al. (1987) reported two studies carried out among the inhabitants of Amsterdam. The first study used large samples of applicants for the position of ticket collector in public transport and used a Verbal Reasoning test (g loading of 0.57), on which first-generation immigrants scored 2.02 SD lower, and a Number Completion test (g loading of 0.51), on which they scored 1.02 SD lower. The second study used a variety of tests, but since groups of high-scoring immigrants and low-scoring immigrants were combined, none of these data are reported here.

Table 5 shows that for the first-generation sample the effect sizes range from, respectively, 0.64 to 1.62 for Turks and Moroccans (weighted average 0.96), 0.74 to 1.49 for Surinamese and Antilleans (weighted average 1.27), 1.02 to 1.62 for the various group (weighted average 1.14), and 0.17 to 0.35 for the mixed group (weighted average 0.29). The ES is +1.00 for South-East Asians, meaning higher scores. For the combination of first- and second-generation samples, the effect sizes range from 1.18 to 1.27 for Turks and Moroccans (weighted average 1.22). For the second-generation samples, the effect sizes range from, respectively, 0.52 to 1.23 for Turks and Moroccans (weighted average 0.77), 0.46 to 0.64 for Surinamese and Antilleans (weighted average 0.55), 0.38 to 0.70 for the

various group (weighted average 0.41), and +0.13 (meaning higher scores) to 0.86 for the mixed group (weighted average 0.05).

When comparing the second generation with the first generation all second-generation groups clearly show smaller effect sizes. For the combination of first- and second-generation Turkish and Moroccan samples, the data show a different trend. This might, however, be explained by the fact that of Hessels' (unpublished doctoral dissertation) test battery four out of six tests are highly *g* loaded. The best support for diminishing group differences is found in studies where groups differ only with regard to generation, respectively, van de Vijver et al. (1993), Helms-Lorenz et al. (2003), and van den Berg (2001). So, these data show that the group differences in *g* are diminishing over time, thereby replicating the findings from the methodologically most sound studies.

Summarizing all studies of nationally and locally representative samples, of complete, specific populations, and of haphazardly collected samples, using effect sizes corrected for language bias, per generation the weighted averages computed on all studies are for the Turks and Moroccans 1.29(1), 1.14(1/2), and 0.83(2), for the Surinamese and Antilleans 1.06(1), 0.60(1/2), and 0.77(2), for the various group 1.13(1) and 0.47(2), for the South-East Asians +0.67 higher, for the mixed group 0.29(1) and 0.08(2); and for the Moluccans 0.40 (2/3). These data again show that group differences in *g* are diminishing over time.

DISCUSSION

The central question addressed in this study is whether cognitive group differences are diminishing over time. The data show substantial improvements for *g*, a remarkable stability of educational differences for younger children, and a clear improvement in educational achievement at the end of primary school. There are no direct data on intergenerational improvements in work proficiency, but indirect data appear suggestive of a trend of temporarily closing gaps.

Educational achievement

Differences in educational achievement between immigrant and majority group children are generally between one-half and one and a half standard deviations, except for the mixed group, for which the differences are smaller. The PRIMA studies show that the effect sizes on numerical tests are smaller than on tests with a high verbal loading, particularly for the Turks and Moroccans. As familiarity with the Dutch language is higher for Surinamese and Antilleans, these results show the strong influence of proficiency in Dutch on achievement test scores.

The effect sizes on the Primary-School Leaving Test are in between the effect sizes of the verbal and numerical subtests of the PRIMA studies, because the Primary School-Leaving Test contains both elements. Over a 10 year period the group differences have remained stable in the lower grades, but for grade 8 the achievement scores have improved by about a third of a standard deviation. The CITO studies, carried out in grade 8, also show a decrease of 0.15 SD between 1987 and 1997. So, these two independent, large-scale studies both show a clear, substantial improvement in educational achievement for immigrant children at the end of primary school.

Work proficiency

Differences in work performance and job-related training results between immigrants and majority group members are generally about a quarter to half a standard deviation, with larger differences for language-related performance aspects. For Turks and Moroccans the differences are larger than for Surinamese and Antilleans. The reported data clearly showed that group differences diminished over time with concern to unemployment rates and mean job level. However, the data from the last study were collected in a time of economic boom and labour shortage. It would be interesting to see whether the findings can be generalized to a period of less prosperity.

g

It is clear that for certain groups the average IQ scores or test scores are underestimates of *g* due to low proficiency in the language of the test and were therefore corrected for bias. The samples in the reported studies are heterogeneous (Cook & Campbell, 1979) in that they employ nationally or locally representative samples, complete, specific samples, and haphazard samples. However, the findings with regard to effect sizes for specific generations are highly consistent. They show that there is a clear rise in the mean level of phenotypic *g*. Group differences in IQ scores and levels of *g* cannot be generalized to other generations. Apart from these increases in *g*, an intergenerational increase in Dutch language proficiency for most groups is also clear. Both improve the speed of integration in Dutch society.

When we set the mean IQ of the Dutch at 100, a rough estimate of the mean level of *g* of first-generation Turkish and Moroccan immigrants is 81; and of first-generation Surinamese and Antillean immigrants it is 84. No empirical data were available for Indonesians, so their mean level of *g* could not be estimated.

For the second-generation Turks and Moroccans a rough estimate of *g* would be 88, and for the second-generation immigrants from the Netherlands Antilles and Surinam it would be 88 also. A rough estimate of *g* for the second- and third-generation Indonesians from the Moluccans is 94. The scores of the mixed group (one Dutch and one immigrant parent) are just a little below the Dutch scores, leading to an estimate of *g* of 99. South-East Asians score much better than the Dutch, but the samples are small, precluding strong conclusions. However, high scores of South-East Asians (Chinese, Japanese, Vietnamese, and Koreans) are well established in other studies (Lynn, 1983), so we conservatively estimate the *g* of South-East Asians in the Netherlands at 105, thereby giving less weight to the small-scale Dutch samples reported in this study and more weight to the large-scale international studies.

Limitations of our study

IQ tests are vehicles for *g* measurement that differ in their ability to measure the construct of *g*. It should not be forgotten that measures of IQ correlate about 0.95 with the battery's own *g* factor, but not with 'true' *g*. We could not estimate the true *g* of the sum scores of the intelligence batteries in the individual studies reported in this paper because to the best of our knowledge there are no studies reporting correlations between all the sum scores from the various test batteries. Jensen (1998, p. 91), however, estimates that IQ tests typically load 0.8–0.9 on true *g*. When we assume that the sum scores of the various intelligence batteries have roughly comparable true *g* loadings, this would mean that the group

differences computed in the present paper underestimate group differences in true *g* by roughly 15%.

Decreasing effect sizes are not necessarily due to a score improvement of immigrant groups over time. Essentially, decreasing group differences only mean that group differences between Dutch and ethnic groups are decreasing. However, there are no signs in the literature of dropping cognitive scores for Dutch children or adults, which increases the plausibility of our conclusions.

Comparison with US findings

Our findings do not show the same pattern as US findings concerning Black/White and Hispanic/White differences. With concern to educational achievement, the mean scores of Black children have risen substantially during the 1970s and the 1980s, in contrast to the achievements of White children. The gains were not restricted to Blacks—Hispanics showed similar rises, albeit not as spectacular. However, the earlier narrowing stopped during the 1980s and started to widen again during the 1990s (Grissmer, Flanagan, & Williamson, 1998; Grissmer, Williamson, Kirby, & Berends, 1998; Hedges & Nowell, 1998). With concern to *g*, it has been suggested that Black/White differences may be diminishing (Thorndike, Hagen, & Sattler, 1986; Vincent, 1991). Lynn (1998), however, finds no support for diminishing group differences and Jensen (1998) and Neisser et al. (1996) conclude that there is no convincing evidence that group differences are diminishing. Ceci, Rosenblum, and Kumpf (1998, p. 292) therefore state that '. . . the racial gap has stubbornly held at 1 SD as long as records have been kept'. We conclude that in the Netherlands the diminishment of group differences in educational achievement is smaller and less general than the substantial diminishment in the US during the 1970s and 1980s, yet it is accompanied by clearly diminishing group differences in *g*, and that this closing of the gap in *g* does not take place in the US.

Consequences for the integration of immigrants

Apart from other causes, such as a substantially lower level of education, the low level of knowledge of Dutch culture, the limited proficiency in Dutch, and discrimination, the low mean level of *g* of immigrants is a clear and important factor that accounts for their lower social–economic status. This conclusion seems justified, because scores on tests of *g* are the best predictors of educational achievement (Jensen, 1980; Neisser et al., 1996) and work proficiency (Schmidt & Hunter, 1998) and because tests seem to predict short-term and long-term criteria about equally well for both Dutch and immigrants (te Nijenhuis & van der Flier, 1999; see Tesser et al., 1999). It also is clear that an intergenerational increase in Dutch language proficiency has taken place for most groups, which also enhances integration. Additional research will be needed to determine what factors cause the strongest impediment and how they are related in terms of cause and effect.

The rise in level of *g* over time most likely indicates that the socio-economic possibilities are improving for this group, but their low mean level of *g* still is a handicap for programs aimed at the integration of immigrants into Western society. It remains to be seen whether new generations of immigrants will gain a position on the labour market that is comparable to that of the Dutch. In this respect, the Indonesians from the Moluccans form an interesting group. Although the status of the small group of third-generation Moluccans on the Dutch labour market is better than that of Surinamese, Antilleans, Turks, and Moroccans, it still has remained behind the status of the ethnic Dutch (Veen &

Robijns, 1994); their estimated mean IQ of 94 (based on one small sample), which is substantially below the Dutch mean, might play a role here.

The substantial increase in *g* over generations is not fully reflected in increasing school achievement. A possible cause may be the increase in the number of so called 'black schools'—schools in which minorities are concentrated. The average teacher in this kind of school is relatively young and inexperienced, and not specially trained for the kind of education these pupils require. Moreover, black schools experience great difficulties in filling vacancies, leading to large numbers of lessons not given or not given in a proper way. This leads to an accumulation of social, psychological, and educational problems (de Jong, 1989; Westerbeek, unpublished doctoral dissertation).

Selective migration as a cause of group differences

Our conclusion that most immigrant groups have a lower mean level of *g* than the mean of the Dutch, South-East Asians have a mean level of g that is higher, and that the mixed group has a mean level of *g* that is virtually identical to that of the Dutch is in line with the finding that group differences in phenotypic *g* are rather the rule than the exception. The reasons for these group differences in mean intelligence still remain unclear, but for certain groups one plausible cause is selective migration. The 1960s and 1970s saw recruitment of workers that came predominantly from the least developed parts of Turkey (Anatolia) and Morocco (Rif mountains) to do unskilled work, and many of these workers had not finished primary school, leading to large-scale functional illiteracy. Lynn and Vanhanen (2002) estimate Turkish IQ at 90, based on the standardization sample ($N = 2277$) of Raven's Standard Progressive Matrices by Sahin and Duzen (1994), and this value is much higher than our estimate for first-generation Turkish immigrants of 81. So, these findings strongly suggests that these Turkish immigrants were not representative of their home countries, which may also be true for Moroccan immigrants.

With concern to Surinamese, the findings are not suggestive of strong selective migration: after the independence of the Dutch colony Surinam, more than half of the population went to the Netherlands, resulting in more Surinamese living in the Netherlands than in Surinam itself. With regard to Netherlands Antilleans, there is proof of selective migration: in the 1960s the Antilleans that went to the Netherlands were predominantly from a small group consisting of outstanding students in secondary education visiting Dutch universities, leading to an average educational level of Antilleans that was higher than that of the Dutch; however, later on many Antilleans with low educational credentials emigrated (Martens & Verweij, 1997).

Indonesians from the Moluccans living in the Netherlands are virtually all soldiers or are from soldier families, and these soldiers were not so much selected on *g*, but more strongly on physical capacities. The estimated mean IQ of 94 (based on one sample of $N = 84$) for Indonesians from the Moluccans is quite close to the estimated value of 89 reported for all of the Indonesian islands by Lynn and Vanhanen (2002, p. 208), and this value is not suggestive of strong selective migration. Taking the problems with Moluccans' integration into consideration, one might hypothesize a slightly lower level of *g*. With concern to the other immigrants from Indonesia, after the Dutch Indies became independent, a large group of those Indonesians working in the Dutch colonial administration—so most likely those with above average education—went to the Netherlands. Also, many of them spoke Dutch fluently, which indicates that many of them went to elite, Dutch-language schools. Moreover, a large group consisted of the offspring from mixed Dutch–Indonesian

marriages, and the findings in this study show that this group has a mean level of *g* that is almost the same of that of the Dutch. All this shows that the immigrants were not representative of their native country. Combined with the fact that Lynn and Vanhanen estimate Indonesian IQ at 89, a plausible conclusion is that the mean IQ of Indonesian immigrants is quite close to the Dutch mean.

With concern to the mixed group, the immigrant partners of Dutch women have been educated better than average for their ethnic group, so they are not representative of their group. The majority of the children from this group have Dutch mothers and hence are being raised according to Dutch norms and have learned Dutch as their mother tongue. Also, the divorce rate among ethnically mixed families is much higher, which means that many children have been raised by their Dutch mother alone.

In the Stams et al. (2000) study, the adopted Korean children score one SD higher than the Dutch average. Frydman and Lynn (1989) report that adopted Korean children score two-thirds of an SD higher than the Belgian average. Also, while Lynn and Vanhanen (2002) estimate IQ in Colombia at 89 and in India at 81, the Colombian and Srilankese adoptees in Stams et al. (2000) score substantially above the Dutch mean. It may be that Western adoptive parents choose the most intelligent and healthy from among the children that are put up for adoption, leading to adopted children not being representative of the populations in their native countries.

Possible causes of improvements in mean level of *g*

One of the most fascinating findings from recent intelligence research is the fact that over generations test scores have gone up, better known as the Flynn effect (Flynn, 1984, 1987). Jensen (1998) suggests that a part of the intergenerational score gains is genuine—people have become cleverer—and that a part is hollow—higher IQ scores that are not reflected in higher real-life problem solving ability. However, in the US the non-genetic factors that are hypothesized to cause the Flynn effect have not narrowed the Black/White IQ gap. With concern to the Netherlands, in a comparison of first- and second-generation immigrants using the WISC, van de Vijver et al. (1993) report that the scores on the verbal tests are 9.8 IQ points higher and that the scores on the performal tests are 12.6 IQ points higher for the second generation. These stronger improvements on the non-verbal tests are in line with the pattern that is consistently found in the work of Flynn. It may be that the factors causing the Flynn effect have not narrowed the IQ gap between groups of Blacks and Whites growing up in the same country, because they had an equally strong influence on each group; however, the gap in *g* between the Dutch and immigrants from underdeveloped countries and their descendants has clearly narrowed and this might possibly be caused by the same factors exerting stronger effects on the immigrants than on the Dutch.

Research shows that apart from educational factors such as years of education (Ceci, 1991; Jensen, 1998) environmental variables of a biological nature in particular influence *g*. Established biological influences on *g* are (i) amount of vitamins and minerals in nutrition (Benton & Roberts, 1988), (ii) amount of animal protein in nutrition (Wachs et al., 1996), (iii) the presence of iodine in nutrition (Bleichrodt & Born, 1994), (iv) the presence of parasitic worms in the body (Watkins & Pollitt, 1997), (v) a high prevalence of consanguineous marriages (Bashi, 1977; Teebi & Farag, 1997). All these educational and biological factors probably impede the first generation of immigrants, born and raised in developing countries, more than the second generation, growing up in post-industrial societies.

CONCLUSION

The findings from this paper show that cognitive differences between the largest immigrant groups and ethnic Dutch are big, but that some of the gaps are diminishing over time. It also shows that some immigrant groups have a higher mean level of *g* than the Dutch, and that the group with one ethnic Dutch and one immigrant parent have a mean level of *g* that is only slightly below the level of the Dutch. It appears that groups with a mean level of *g* that is substantially below that of the Dutch have difficulties integrating, and that groups with a mean level of *g* that is comparable to or even higher than that of the Dutch experience for fewer problems. As for many West-European countries immigrants come from Third-World countries, including former colonies, these Dutch findings probably may be generalized to these West-European countries. Diminishing group differences over time enhance the process of integration into Western society.

ACKNOWLEDGEMENTS

Part of this research was supported by an NWO, the Dutch Organization for Scientific Research, grant awarded to the first author. The authors wish to thank Dr Henny Uiterwijk for providing the 1997 CITO data on school achievement, Dr Paul Tesser for providing the data of the PRIMA studies, Dr Remko van den Berg and Dr Nico Bleichrodt for providing the MCT-M data, Dr Katrien van de Vijfeijken for providing the Draw-A-Man data, and Dr Michelle Helms-Lorenz and Dr Fons van de Vijver for providing the data from Helms-Lorenz, van de Vijver, and Poortinga (2003). The authors thank Linda Gottfredson and an anonymous reviewer for their detailed and constructive criticism of an earlier version of this paper.

REFERENCES

Bashi, J. (1977). Effects of inbreeding on cognitive performance of Israeli Arab children. *Nature, 266*, 440–442.

Benton, D., & Roberts, G. (1988). Effect of vitamin and mineral supplementation on intelligence of a sample of schoolchildren. *Lancet, 333*, 140–144.

Bleichrodt, N., & Born, M. Ph. (1994). A meta-analysis of research on iodine and its relationship to cognitive development. In J. B. Stanbury (Ed.), *The damaged brain and iodine deficiency* (pp. 195–200). New York: Cogizant Communication.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Ceci, S. J. (1991). How much does schooling influence intellectual development and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*, 703–722.

Ceci, S. J., Rosenblum, T. B., & Kumpf, M. (1998). The shrinking gap between high- and low-scoring groups: Current trends and possible causes. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 287–302). Washington, DC: American Psychological Association.

Centraal Bureau voor de Statistiek [Dutch Central Bureau for Statistics]. (2000). *Allochtonen in Nederland 1999* [*Minorities in The Netherlands 2000*]. Den Haag: Centraal Bureau voor de Statistiek.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton.

Dagevos, J. M. (1996). Allochtonen in de functiestructuur. Individuele kenmerken, groepsgebonden oriëntaties, etnische functies en ongelijke kansen [Immigrants in the job hierachy. Individual characteristics, group directed orientations, ethnic jobs and unequal opportunities]. *Tijdschrift voor Arbeidsvraagstukken, 12*, 4–17.

Dagevos, J. M. (2001a). Interne promotiemogelijkheden van etnische minderheden. De betekenis van menselijk kapitaal, baankenmerken en ongelijke behandeling [Internal promotion possibilities of immigrants. The role of human capital, job characteristics and unequal treatment]. *Tijdschrift voor Arbeidsvraagstukken, 17*, 380–389.

Dagevos, J. M. (2001b). *Rapportage minderheden 2001. Deel II Meer werk* [*Report on minorities 2001. Part II More work*]. Den Haag: Sociaal en Cultureel Planbureau.

Dagevos, J. M., & Veenman, J. (1996). Werkloosheid en het voortduren van achterstand [Unemployment and continuation of disadvantage]. In J. Veenman (Ed.), *Keren de kansen? De tweede-generatie allochtonen in Nederland* [*Do their chances improve? The second generation immigrants in the Netherlands*] (pp. 39–62). Assen, the Netherlands: van Gorcum.

de Jong, M. J. (1987). *Herkomst, kennis en kansen: Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs* [*Origin, knowledge, and opportunities: Immigrant and majority group pupils during the transition from primary to secondary education*]. Lisse: Swets.

de Jong, M. J. (1989). Toenemende apartheid in het Nederlandse onderwijs: Oorzaken, gevolgen en een mogelijke remedie [Increasing separation in Dutch education: Causes, consequences, and a possible remedy]. *Pedagogische Studiën, 88*(2), 61–73.

de Jong, M. J., & van Batenburg, Th. A. (1984). Etnische herkomst, intelligentie en schoolkeuzeadvies [Ethnic origin, intelligence, and school choice advice]. *Pedagogische Studiën, 61*, 362–371.

de Vries, S., & Pettigrew, T. F. (1998). Effects of ethnic diversity: The position of minority workers in two Dutch organizations. *Journal of Applied Social Psychology, 28*, 1503–1530.

Driessen, G. (1991). Discrepanties tussen toetsresultaten en doorstroomniveau. Positieve discriminatie bij de overgang basisonderwijs—voortgezet onderwijs? [Discrepancies between test results and transfer level. Positive discrimination during the transition from primary education to secondary education]. *Pedagogische Studiën, 68*, 27–35.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29–51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Frydman, M., & Lynn, R. (1989). The intelligence of Korean children adopted in Belgium. *Personality and Individual Differences, 10*, 1323–1325.

Grissmer, D., Flanagan, A., & Williamson, S. (1998). Why did the Black–White score gap narrow in the 1970s and 1980s? In C. Jencks, & M. Phillips (Eds.), *The Black–White test score gap* (pp. 182–226). Washington, DC: Brookings Institution Press.

Grissmer, D. W., Williamson, S., Kirby, S. N., & Berends, M. (1998). Exploring the rapid rise in black achievement scores in the United States (1970–1990). In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 251–285). Washington, DC: American Psychological Association.

Harrison, P. L., & Schock, H. H. (1994). Draw-a-Figure Test. In R. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 361–362). New York: Macmillan.

Hedges, L. V., & Nowell, A. (1998). Black–White test score convergence since 1965. In C. Jencks, & M. Phillips (Eds.), *The Black–White test score gap* (pp. 149–181). Washington, DC: Brookings Institution Press.

Helms-Lorenz, M., van de Vijver, F. J. R., & Poortinga, Y. H. (2003). Cross-cultural differences in cognitive performance and Spearman's hypothesis: g or c? *Intelligence, 31*, 9–29.

Jensen, A. R. (1980). *Bias in mental testing*. London: Methuen.

Jensen, A. R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences, 8*, 193–219.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* London: Praeger.

Koeslag, M., & Dronkers, J. (1994). Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs [Over-advising and the school careers of immigrants and majority group pupils in secondary education]. *Tijdschrift voor Onderwijsresearch, 19*, 240–258.

Lopez, E. C. (1997). The cognitive assessment of limited English proficient and bilingual children. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 503–516). New York: Guilford.

Lubinski, D., & Dawis, R. V. (1992). Aptitudes, skills, and proficiencies. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organisational psychology* (2nd ed., Vol. 3). Palo Alto: Consulting Psychologists Press.

Lynn, R. (1983). IQ in Japan and the United States. *Nature, 306*, 291–292.

Lynn, R. (1998). Has the black–white intelligence difference in the United States been narrowing over time? *Personality and Individual Differences, 25*, 999–1002.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations.* Westport, CT: Praeger.

Martens, E. P., & Verweij, A. O. (1997). *Antillianen en Arubanen in Nederland: Kerncijfers 1996* [*Netherlands Antilleans and Arubans in the Netherlands: Essential figures 1996*]. Rotterdam, the Netherlands: Instituut voor Sociologisch-Economisch Onderzoek.

Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 277–101.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Ogbu, J. U. (1994). Culture and intelligence. In R. J. Sternberg (Ed.), *Ecyclopedia of human intelligence* (pp. 328–338). New York: Macmillan.

Pennock-Román, M. (1992). Interpreting test performance in selective admissions for Hispanic students. In K. F. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 99–135). Washington, DC: American Psychological Association.

Resing, W. C. M., Bleichrodt, N., & Drenth, P. J. D. (1986). Het gebruik van de RAKIT bij allochtoon etnische groepen [Use of the RAKIT for assessment of immigrants]. *Nederlands Tijdschrift voor de Psychologie, 41*, 179–188.

Sahin, N., & Duzen, E. (1994). Turkish standardisation of Raven's SPM. In *Proceedings of the 23rd International Congress of Applied Psychology*, Madrid.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities as predictors of work behaviours: A European contribution and comparison with American findings. *Personnel Psychology, 56*, 573–605.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Snijders, J. Th., Tellegen, P. J., & Laros, J. A. (1988). *Snijders–Oomen niet-verbale intelligentietest: S.O.N.-R 5.5-17. Verantwoording en handleiding* [*Snijders–Oomen nonverbal intelligence test S.O.N. 5.5-17. Justification and manual*]. Groningen: Wolters-Noordhoff.

Spijkerman, R. M. H. (2000). *De positie van allochtonen en autochtonen in het bedrijfsleven en bij een deel van de overheid* [*The position of immigrants and majority group members in industry and part of the governmental organizations*]. Den Haag: Elsevier Bedrijfsinformatie.

Stams, G.-J. J. M., Juffers, F., Rispens, J., & Hoksbergen, R. A. C. (2000). The development and adjustment of 7-year-old children adopted in infancy. *Journal of Child Psychology and Psychiatry, 41*, 1025–1037.

Teebi, S. T., & Farag, T. I. (1997). *Genetic disorders among Arab populations*. Oxford: Oxford University Press.

Tellegen, P. J. (2002). De WISC-III: Een illusie armer [The WISC III: One illusion less]. *De Psycholoog, 37*, 607–610.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B., & Laros, J. (1998). *Snijders–Oomen niet-verbale intelligentietest: SON-R 2.5-7. Handleiding en verantwoording* [*Snijders–Oomen nonverbal intelligence test: SON-R 2.5-7. Manual and justification*]. Lisse: Swets.

te Nijenhuis, J. (2001). Enige theoretische en empirische problemen bij intelligentiemeting met alternatieve tests: Een reactie op Van den Berg en Bleichrodt (2000) [Some theoretical and empirical problems when testing with alternative intelligence tests: A reaction to van den Berg and Bleichrodt (2000)]. *Nederlands Tijdschrift voor de Psychologie, 56*, 10–15.

te Nijenhuis, J., Evers, A., & Mur, J. P. (2000). Validity of the Differential Aptitude Test for immigrant children. *Educational Psychology, 20*, 99–115.

te Nijenhuis, J., Tolboom, E. A. M., Resing, W. C. M., & Bleichrodt, N. (in press). Does cultural background influence the intellectual performance of children from immigrant groups?: Validity

of the RAKIT intelligence test for immigrant children. *European Journal of Psychological Assessment*.

te Nijenhuis, J., & van der Flier, H. (1997). Comparability of GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology, 82*, 675–687.

te Nijenhuis, J., & van der Flier, H. (1999). Bias research in the Netherlands: Review and implications. *European Journal of Psychological Assessment, 15*, 165–175.

te Nijenhuis, J., & van der Flier, H. (2000). Differential prediction of immigrant versus majority group training performance using cognitive ability and personality measures? *International Journal of Selection and Assessment, 8*, 54–60.

te Nijenhuis, J., & van der Flier, H. (2001). Group differences in mean intelligence for the Dutch and third world immigrants. *Journal of Biosocial Science, 33*, 469–475.

te Nijenhuis, J., & van der Flier, H. (2003). Immigrant–majority group differences in cognitive performance: Jensen effects, cultural effects, or both? *Intelligence, 31*, 443–459.

Tesser, P. T. M., & Iedema, J. (2001). *Rapportage Minderheden 2001: Vorderingen op school*. [*Report on ethnic minorities 2001: Achievement in school*]. Den Haag: Sociaal en Cultureel Planbureau.

Tesser, P. T. M., Merens, J. G. F., & van Praag, C. (1999). *Rapportage minderheden 1999: Positie in het onderwijs en op de arbeidsmarkt* [*Report on minorities 1999: Position in education and on the labor market*]. Rijswijk, the Netherlands: Sociaal en Cultureel Planbureau.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet intelligence scale: Fourth edition (Technical manual)*. Chicago, IL: Riverside.

Uiterwijk, H. (1994). *Eindtoets Basisonderwijs. De bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen* [*Suitability of the Primary- School Leaving Test for students from non-Dutch backgrounds*]. Arnhem, the Netherlands: CITO.

van den Berg, R. H. (2001). *Psychologisch onderzoek in een multiculturele samenleving: Psychologische tests, interview- en functioneringsbeoordelingen* [*Psychological research in a multicultural society: Psychological tests, interview, and job proficiency assesment*]. Amsterdam: Stichting NOA.

van den Berg, R. H., & Bleichrodt, N. (2000). Intelligentiemeting bij kandidaten met verschillende culturele achtergronden: de Multiculturele Capaciteiten Test (MCT-M) [Intelligence measurement of candidates with varying cultural backgrounds: the Multicultural Capacities Test (MCT-M]. *Nederlands Tijdschrift voor de Psychologie, 55*, 134–147.

van de Vijfeijken, K., & Vedder, P. (2000). De cultuurgevoeligheid van de mens-teken-test [The sensitivity to culture of the Draw-a-Man test]. *Pedagogische Studiën, 77*, 166–172.

van de Vijfeijken, K., Vedder, P., & Kook, H. (1997). Menstekening en cognitieve ontwikkeling op Curaçao [Draw-a-Man and cognitive development on Curaçao]. *Nederlands Tijdschrift voor Opvoeding, Vorming en Onderwijs, 13*, 262–274.

van de Vijver, F., Willemse, G., & van de Rijt, B. (1993). Het testen van cognitieve vaardigheden van allochtone leerlingen [Testing of cognitive skills of immigrant pupils]. *De Psycholoog, 48*, 152–159.

van Gastel, J. H. M., Bleichrodt, N., & van Heelsum, A. J. (1994). Het gelijk van de praktijk: Loopbaanontwikkeling en werkbeleving van allochtonen en autochtonen [Practice in the right: Career development and work satisfaction of immigrants and majority group members]. In N. Bleichrodt (Ed.), *Gelijke Kansen op Werk: Evenredige vertegenwoordiging van allochtonen en autochtonen in 2000* [*Equal opportunities for work: Equal proportions of immigrants and majority group members in 2000*] (pp. 53–109). Amsterdam: Stichting PION.

van Leest, P. F. (1997). *Persoonlijkheidsmeting bij allochtonen: De gebruikswaarde van persoonlijkheidsvragenlijsten bij selectie van allochtonen* [*Assessment of personality for ethnic minorities: The usefulness of personality questionnaires for the selection of immigrants*]. Lisse: Swets.

van Leest, P. F., & Dutman, A. (1995). Psychologische selectie voor de Turken-en Marokkanenpool: Onderzoek naar de voorspellende waarde [Psychological selection for the Turks and Moroccans pool: A study into predictive validity]. *Het Tijdschrift voor de Politie, 57*(4), 14–17.

van Rooijen, E. (1992). *Het gebruik van psychologische tests bij allochtonen in het kader van selektie* [*The use of psychological tests for assessment of immigrants for selection*]. Paper presented at WAOP conference, University of Groningen.

Veen, A., & Robijns, M. (1994). Schoolloopbanen van Molukse leerlingen in het voortgezet onderwijs [School careers of Moluccan pupils in secondary education]. *Tijdschrift voor Onderwijsresearch, 19*, 143–159.

Verhaeghe, P., Schittekatte, M., Vermeir, G., & Bosmans, M. (2002). *Dé nieuwe Nederlandstalige intelligentietest: De WISC-III* [*The new Dutch-language intelligence test: The WISC-III*]. Press release, University of Ghent.

Verouden, G., Ross, F., Stet, A., & Scheele, J. (1987). *Psychologische selektie van etnische minderheden: Verslag van een onderzoek naar voor allochtonen bruikbare testmethoden ten behoeve van de gemeente Amsterdam* [*Psychological selection of ethnic minorities: Report of a study into testing methods that can be used for immigrants for the municipality of Amsterdam*]. Amsterdam: Gemeentelijke Geneeskundige en Gezondheidsdienst.

Vincent, K. R. (1991). Black/White IQ differences: Does age make the difference? *Journal of Clinical Psychology, 47*, 266–270.

Wachs, T. D., McCabe, G., Mousa, W., Yunis, F., Kirksey, A., Galal, O., Harrison, G., & Jerome, N. (1996). Cognitive performance of Egyptian adults as a function of nutritional intake and sociodemographic factors. *Intelligence, 22*, 129–154.

Watkins, W. E., & Pollitt, E. (1997). 'Stupidity or worms'. Do intestinal worms impair mental performance? *Psychological Bulletin, 121*, 171–191.

